PEABODY JOURNAL OF EDUCATION, 77(2), 64–84

Curriculum-Based Measurement: Describing Competence, Enhancing Outcomes, Evaluating Treatment Effects, and Identifying Treatment Nonresponders

Lynn S. Fuchs and Douglas Fuchs
Peabody College
Vanderbilt University

The authors summarize research on curriculum-based measurement (CBM) within four strands. They provide an overview of studies demonstrating the psychometric tenability of CBM. They discuss the body of work showing how teachers can use CBM to inform instructional planning. They examine CBM's potential use in evaluating treatment effects. Finally, they summarize work on CBM for the purpose of identifying children who fail to profit from otherwise effective instruction.

Curriculum-based measurement (CBM) is a set of methods for indexing academic competence and progress. The goal in developing CBM (Deno, 1985) was to establish a measurement system that teachers could use efficiently to (a) obtain accurate, meaningful information with which to index standing and growth; (b) answer questions about the effectiveness of programs in producing academic learning; and (c) plan better instructional programs. To accomplish this goal, a systematic program of research, conceptualized as a 3 × 3 matrix (Deno & Fuchs, 1987), was undertaken. The rows of this matrix specified questions for developing a measurement system (what to measure, how to measure, and how to use the resulting database); the columns provided criteria that answer those

Requests for reprints should be sent to Lynn S. Fuchs and Douglas Fuchs, Peabody College, Box 330, Nashville, TN 37203.

questions (technical adequacy, treatment validity, and feasibility). A 20-year research program, undertaken by independent investigators at multiple sites, has addressed the cells in this matrix, with multiple studies for reading, spelling, mathematics, and written expression.

In each academic area, CBM integrates key concepts from traditional measurement theory and the conventions of classroom-based observational methodology to forge an innovative approach to assessment. As with traditional measurement, every assessment samples a relatively broad range of skills by sampling each dimension of the annual curriculum on each weekly test; consequently, each repeated measurement is an alternate form, of equivalent difficulty, assessing the same constructs. This sampling strategy differs markedly from typical classroom-based assessment methods, in which teachers assess mastery on a single skill, and after mastery is demonstrated, move on to a different, presumably more difficult skill (L. S. Fuchs & Deno, 1991; L. S. Fuchs & Fuchs, 1999). CBM also relies on a traditional psychometric framework by incorporating conventional notions of reliability and validity so the standardized test administration and scoring methods yield accurate and meaningful information.

By sampling broadly with standardized administration and scoring procedures, the CBM score can be viewed as a *performance indicator.* That is, the CBM score produces a broad dispersion of scores across individuals of the same age, with rank orderings that correspond to important external criteria; it represents an individual's global level of competence in the domain. Practitioners can use this performance indicator to identify discrepancies in performance levels between individuals and peer groups, which helps inform decisions about the need for special services or the point at which decertification and reintegration of students with disabilities occur.

At the same time, however, CBM departs from conventional psychometric applications by integrating the concepts of standardized measurement and traditional reliability and validity with key features from classroom-based observational methodology: repeated performance sampling, fixed time recording, graphic displays of time-series data, and qualitative descriptions of student performance. Reliance on these classroom-based observational methods permits slope estimates for different time periods and alternative interventions for the same individual; this creates the necessary database for describing growth and testing the effects of different treatments for a given student. Also, when combined with prescriptive decision rules, these time-series analytic methods result in better instruction and learning. Teachers raise goals more often and develop higher expectations (L. S. Fuchs, Fuchs, & Hamlett, 1989a), adapt their instruc-

tion more frequently (L. S. Fuchs, Fuchs, & Hamlett, 1989b), they effect better student learning (L. S. Fuchs, Fuchs, Hamlett, & Stecker, 1991).

In addition, because each assessment simultaneously samples the multiple skills embedded in the annual curriculum, CBM can yield qualitative descriptions of student performance to supplement graphed, quantitative analyses of CBM total scores. These diagnostic profiles demonstrate reliability and validity (see, e.g., L. S. Fuchs, Fuchs, Hamlett, & Allinder, 1989; L. S. Fuchs, Fuchs, Hamlett, Thompson, et al., 1994), offer the advantage of being rooted in the local curriculum, provide a framework for improving student programs, and result in teachers planning more varied, specific, and responsive instruction to meet individual student needs (L. S. Fuchs, Fuchs, Hamlett, & Allinder, 1991a).

CBM bridges traditional psychometric and classroom-based observational assessment paradigms to forge an innovative approach to measurement. Through this bridging of frameworks, CBM simultaneously yields information about standing as well as change and about global competence as well as skill-by-skill mastery. CBM can, therefore, be used to answer questions about interindividual differences (e.g., How different is Henry's academic level and growth from that of other students in the class, school, or district?); intra-individual improvement (e.g., How successful is an adapted regular classroom in producing better academic growth for Henry?); and how to strengthen individual student programs (e.g., On which skills in the annual curriculum does Henry require instruction?).

In this article, we summarize research on CBM within four strands: (a) We provide an overview of studies demonstrating the psychometric tenability of CBM. (b) We discuss the body of work demonstrating how teachers can use CBM to inform instructional planning and effect superior learning. (c) We examine CBM's potential use in evaluating treatment effects. (d) We summarize work on CBM for the purpose of identifying children who fail to profit from otherwise effective instruction; that is, the group of students referred to alternately as treatment *resisters* or *nonresponders*.

Psychometric Tenability of CBM

We illustrate the psychometric strengths of CBM by briefly summarizing information in the area of reading. In mathematics, spelling, and written expression, similar data exist; however, the vast majority of independent replications do occur in reading. We organize our discussion in terms of technical features necessary for describing student competence at one time as opposed to the features needed to model student growth over time.

Describing Student Competence at a Specific Time

To provide the basis for sound decision making about a student's performance level, an assessment score (or an average across several scores) must provide an accurate and meaningful estimate of competence; therefore, traditional psychometric methods for investigating technical adequacy apply. To achieve these traditional psychometric criteria, assessment methods typically sample behavior broadly, rely on standardized administration and scoring procedures, and accordingly are viewed as performance indicators. This is the case for CBM, which illustrates how a classroom-based assessment method can achieve traditional psychometric standards.

In reading, two CBM measures are the number of words read aloud from text in 1 min and the number of correct replacements restored to text from which every seventh word has been deleted in 2.5 min. Each measure demonstrates strong criterion validity with respect to widely used, commercial reading tests (L. S. Fuchs & Fuchs, 1992; Marston, 1989), to informal reading measures involving question answering, close completion, recall of passages (L. S. Fuchs & Fuchs, 1992; L. S. Fuchs, Fuchs, & Maxwell, 1988), and to teachers' judgments of reading competence (L. S. Fuchs & Fuchs, 1992; L. S. Fuchs, Fuchs, & Deno, 1982). In addition, evidence suggests adequate (a) predictive validity (e.g., D. Fuchs, Fuchs, Thompson, et al., 2001), and construct validity (e.g., Shinn, Good, Knutson, Tilly, & Collins, 1992), (b) discriminative validity with respect to special education status (Deno, Mirkin, & Chiang, 1982; Shinn, Tindal, Spira, & Marston, 1987) and grade (Deno, 1985; L. S. Fuchs & Deno, 1992; L. S. Fuchs, Fuchs, Hamlett, Walz, & Germann, 1993), (c) stability (L. S. Fuchs, Deno, & Marston, 1983; L. S. Fuchs & Fuchs, 1992), and (d) interscorer agreement (L. S. Fuchs, Fuchs, Hamlett, & Ferguson, 1992; Marston & Deno, 1981).

In the most recent study of CBM's psychometric features, Hosp and Fuchs (2001) had 74 first graders, 81 second graders, 79 third graders, and 75 fourth graders read two CBM grade-appropriate passages; they had the students complete the Woodcock Reading Mastery Tests and the word identification, word attack, and passage comprehension subtests; and they re-administered the CBM passages 2–3 weeks later to a subsample of the participants (i.e., 29, 30, 30, and 30 at the four grades, respectively). At these respective grades, criterion validity between CBM and word identification was .89, .88, .89, and .79; between CBM and word attack, .70, .82, .82, and .74; and between CBM and passage comprehension, .76, .83, .83, and .83. Test–retest reliability was .96, .97, .93, and .93. In another study (Fuchs & Hamlett, 1997), predictive validity with respect to a commercial

group-administered achievement test, the TerraNova, was .86 over 12 weeks (January–April) as documented among 324 first-grade students; .79 over 30 weeks (October–May) as documented among 87 second-grade students; and .72 over 30 weeks (October–May) as documented among 95 third-grade students.

Modeling Academic Growth

Performance indicators, which are commonly associated with formal testing, are also important to the classroom-based assessment methods, such as CBM, for modeling growth. This is the case for two reasons: (a) Performance indicators provide a broad range of scores required for manifesting change over time and (b) the traditional standards of psychometric adequacy on which performance indicators are based provide necessary evidence for presuming that differences between an individual's data points represent meaningful change.

Unfortunately, however, these traditional psychometric criteria are insufficient evidence that a measure can adequately depict growth. The modeling of longitudinal data to describe growth has been discussed for some time in the biological and statistical literatures and has been applied more recently in psychology (e.g., Willett, Ayoub, & Robinson, 1991) and education. As discussed by Francis, Shaywitz, Stuebing, Shaywitz, and Fletcher (1994), instruments for longitudinally modeling individual change must demonstrate certain technical features, which CBM demonstrates.

First, the instrument must provide equal scaling of individuals throughout the range of behavior measured over time (i.e., produce data with interval scale properties, free from ceiling or floor effects). With CBM, a common test framework is administered to children within a fixed age range, making it possible to judge performance over an academic year on the same raw score metric. When performance is measured on the appropriate instructional level of the curriculum, floor and ceiling effects do not occur. Second, the construct and the difficulty level measured over time must remain constant. CBM taps constructs that are qualitatively constant over an academic year, for which the difficulty level remains the same. The third technical requirement for the modeling of growth is that a sufficient number of alternate forms be available to obtain accurate estimates of change parameters. With CBM, alternate passages of constant difficulty are available (9–30 passages per grade, depending on source), and research (L. S. Fuchs, 1993) suggests that approximately 7 data points are adequate for fitting data to a model.

Current techniques for measuring change reconceptualize growth as a continuous, rather than an incremental, process. The goal is to describe trajectories, or continuous time-dependent curves, which reflect the change process. An initial step in such a process is to formulate a change model at the individual level. Examination of individual and group time-series CBM data provides the basis for an empirical approximation of the shape of CBM growth curves (Francis et al., 1994). L. S. Fuchs, et al. (1993), for example, examined students' academic growth rates when CBM was conducted for one school year in students' grade-appropriate curriculum level. For many students on each CBM measure, a linear relationship adequately modeled student progress within one academic year. When significant quadratic terms occurred (for 0%–21% of students), growth was almost consistently described by a negatively accelerating pattern, in which student performance continues to improve over the course of a year, but the amount of that progress gradually decreases. As suggested in cross-sectional data, this negatively accelerating pattern may also characterize growth across academic years. These findings, in combination with corroborating evidence (Good, Deno, & Fuchs, 1995; Good & Shinn, 1990), support a conceptualization of annual CBM growth characterized by a linear relationship, where slope is a primary parameter describing change. Consequently, CBM appears to be a tenable measurement tool for modeling academic growth.

Using CBM to Enhance Instructional Planning and Student Learning

A well-established, long-standing research program documents how CBM can help teachers plan better instruction and effect superior achievement. Several investigators have examined the effects of alternative data-utilization strategies, as well as CBM's overall contribution to instructional planning and student learning, not only in regular educationbut in special education as well.

Effects of Alternative Data-Utilization Strategies

CBM has been shown to enhance teacher planning and student learning by helping teachers set ambitious student goals, by assisting teachers in determining when instructional adaptations are necessary to prompt better student growth, and by providing ideas for potentially effective teaching adjustments.

Setting ambitious goals. Studies illustrate how teachers may use CBM to help them establish ambitious goals, which result in accelerated student learning. Fuchs, Fuchs, and Hamlett (1989a), for example, explored the contribution of goal-raising guidelines within CBM decision-making rules. Teachers were assigned randomly to and participated in one of three treatments for 15 weeks in mathematics: no CBM, CBM without a goal-raising rule, and CBM with a goal-raising rule. The goal-raising rule required teachers to increase goals whenever the student's actual rate of growth, represented by an ordinary least squares regression through at least 7 CBM scores, was greater than the growth rate anticipated by the teacher. Teachers in the CBM goal-raising condition raised goals more frequently (for 15 of 30 students) than did teachers in the nongoal-raising conditions (for 1 of 30 students); moreover, concurrent with teachers' goal raising was differential student achievement on pre- and poststandardized achievement tests: The effect size comparing the pre- and post-change of the two CBM conditions (i.e., with and without the goal-raising rule) was .52 standard deviations. Using CBM to monitor the appropriateness of instructional goals and adjust goals upward whenever possible represents one means by which teachers can use CBM to enhance their instructional planning.

Identifying when to revise the instructional program. A second way in which CBM can be used to enhance instructional decision making is to assess the adequacy of student progress and determine whether, and if so when, instructional adaptation is necessary. When the actual growth rate (ordinary least squares regression line through at least 7 CBM scores) is less than the expected growth rate (slope of the goal line), the teacher modifies the instructional program to promote stronger learning. L. S. Fuchs, et al. (1989b) estimated the contribution of this CBM decision-making strategy with 29 special educators who implemented CBM for 15 school weeks with 53 students with mild to moderate disabilities. Teachers in a CBM-measurement only group measured students' reading growth as required but did not use the assessment information to structure students' reading programs. Teachers in the CBM-change the program decision-rule group measured student performance and used the assessment information to determine when to introduce programmatic adaptations to enhance growth rates. Results indicated that although teachers in both groups measured student performance, important differences were associated with the use of the change the program decision rule. As indicated on the Stanford Achievement Test–reading comprehension subtest, students in the change the program decision-rule group

achieved better than did a no-CBM control group (effect size = .72), whereas the measurement only CBM group did not (effect size = .36). The slopes of the two CBM treatment groups were significantly different, favoring the achievement of the change the program group (effect size = .86). As suggested by these findings and the results of other researchers (e.g., Stecker & Fuchs, 1999; Wesson, Skiba, Sevcik, King, & Deno, 1984), collecting CBM data, in and of itself, exerts only a small effect on student learning. To enhance student outcomes in important ways, teachers need to use the CBM data instrumentally to build effective programs for difficult-to-teach students.

Using CBM diagnostic information to plan instruction. To help teachers determine when adjustments are required in students' programs, thus to identify when goal increases are warranted, the CBM total scores are used. In addition, by inspecting the graph of performance indicators over time, teachers may formulate ideas for potentially effective instructional adaptations. For example, a flat or decelerating slope might generate hypotheses about lack of maintenance of previously learned material or about motivational problems; nevertheless, to obtain rich descriptions of student performance, alternative ways of summarizing and describing student performance are necessary. Because CBM assesses performance on the year's curriculum at each testing, rich descriptions of strengths and weaknesses in the curriculum can be generated.

During the 1987–1988 academic year, we investigated the contribution of these CBM diagnostic profiles in math (L. S. Fuchs, et al., 1991), reading (L. S. Fuchs, Fuchs, & Hamlett, 1989c), and spelling (L. S. Fuchs, Fuchs, Hamlett, & Allinder, 1991a). In each investigation, teachers were assigned randomly to one of three conditions: no CBM, CBM with goal-raising and change-the-program decision rules, and CBM with goal-raising and change-the-program decision rules along with CBM diagnostic profiles. In all three studies, teachers in the diagnostic profile treatment group generated instructional plans that were more varied and more responsive to individuals' learning needs. Moreover, they effected better student learning as measured on change between pre- and posttest performance on global measures of achievement. Effect sizes associated with the CBM diagnostic profile groups ranged from .65 to 1.23. This series of studies demonstrated how structured, well-organized CBM information about students' strengths and difficulties in the curriculum helps teachers build better programs and effect greater learning.

More recently, we explored additional methods for deriving additional diagnostic information from the CBM reading aloud score. The first step

in this research program was to identify CBM cut-points, by grade level, at which students required decoding versus fluency versus comprehension intervention. To accomplish this, Hosp and Fuchs (2001) had 74 first graders, 81 second graders, 79 third graders, and 75 fourth graders read two CBM grade-appropriate passages; they had students complete the Woodcock Reading Mastery Tests word identification, word attack, and passage comprehension subtests; and they had students complete the Decoding Skills Battery (Fuchs, Hosp, Fuchs, & Hamlett, 2001b), which assesses student mastery of 10 decoding skills. Using discriminant function analyses on CBM, the Decoding Skills Battery, and Woodcock performance, we established the cut-points. Teachers then used CBM in conjunction with the Decoding Skills Battery to plan instruction in the following way. Every 3 weeks, students were categorized, using their most recent CBM score, into decoding-, fluency-, or comprehension-building activities. For students classified as requiring decoding work, the Decoding Skills Battery was administered (1–5 min using basals and ceilings) to categorize students further into the decoding skills on which instruction should occur. Fuchs, Hosp, Fuchs, and Hamlett (2001a) are investigating the effects of this CBM diagnostic feedback among 28 second-grade teachers who were assigned randomly to four groups: control, instructional consultation, instructional consultation with CBM, and instructional consultation with CBM plus diagnostic feedback. Effects are being assessed on instructional planning and student learning.

CBM's Overall Contribution to Teacher Planning and Student Achievement in Regular Classrooms

As described by Marston and Magnusson (1988), the Minneapolis Public Schools incorporated CBM prereferral assessment within its eligibility assessment process. Over a period of 6 weeks, interventions were implemented and ongoing CBM data were collected to assess the extent to which students' academic needs could be addressed in the regular classroom when instructional adaptations had been introduced. Only pupils whose performance did not improve as a function of these adaptations were identified for special education services. Marston and Magnusson (1988) reported that among students initially referred, 25–45% were deemed eligible for special education after CBM prereferral assessment. This figure is dramatically lower than the estimate reported by Algozzine, Christenson, and Ysseldyke (1982), in which 90% of referred students were subsequently identified for special education using conventional assessment procedures.

During the 1992–1993 school year, L. S. Fuchs, Fuchs, Hamlett, Phillips,

and Karns (1995) further studied the viability of CBM prereferral assessment. We randomly assigned general educators to two treatments. In both treatments, teachers implemented CBM in mathematics with all students in their classes beginning in September. In addition, to facilitate the link between CBM and instruction, teachers in both conditions incorporated a demonstrably effective structured form of peer-assisted learning (e.g., D. Fuchs, Fuchs, Mathes, & Simmons, 1997; L. S. Fuchs, Fuchs, Hamlett, Phillips, et al., 1997).

This combination of CBM and peer-assisted learning strategies therefore represented the baseline treatment in this study, over which individual instructional adaptations were layered. Beginning in November, the bimonthly CBM class reports identified up to two target students per class whose CBM progress was inadequate (i.e., low level combined with low slope, relative to classmates). For these students, teachers (a) formulated an adaptation before the next 2-week report, (b) implemented that adaptation at least four times in the upcoming 2 weeks, and (c) modified previous adaptations to enhance progress when CBM identified the same student over multiple reports.

Results demonstrated that when general educators are specifically prompted with CBM and supported to engage in instructional adaptation, they do so with respectable fidelity. Across three to six 2-week adaptation cycles, teachers ignored requests for adaptations only infrequently; they often implemented multiple strategies concurrently to address the problems of target students; and some teachers modified student programs repeatedly in a variety of ways in an attempt to boost progress. Moreover, teacher reliance on individual adaptations appeared to prompt changes in their thinking about differentiating instructional plans. Compared to teachers in the baseline treatment, those in the adaptations treatment reported (a) more modifications in their goals and strategies for poorly progressing students, (b) a greater variety of skills taught, (c) selective reteaching of lessons more frequently, and (d) more frequent deviation from the teacher's manual for selected students.

Findings were not uniformly positive, however. Despite many focused attempts to enhance learning, some children proved unresponsive to regular classroom adaptations. Two brief cases illustrate this differential responsiveness. Over a 12-week period, a fourth-grade teacher implemented a rich set of adaptations, relying on basic facts drills, motivational workcharts and contracts, and manipulatives. The target student, who exhibited a CBM slope of .21 digits per week when identified for adaptation, responded well to these modifications to the regular classroom and completed the school year with a slope of .63 digits per week—the average slope for the class. This success contrasts with the experience of a third-

grade teacher who also implemented a large number of adaptations including drilling basic facts, sliding back to second-grade material, implementing a motivational workchart, and using money to work on conceptual underpinnings. Despite this teacher's similar level of effort to modify regular classroom instruction, her target student demonstrated little improvement in growth rate: He ended the year with a relatively low slope of .28 digits per week, which was similar to his slope at the time he was identified for adaptation and which was considerably lower than his classmates' average slope of .98 digits per week.

Three of our 10 teachers effected substantial improvement for target students. This suggests that, with the assistance of rich assessment information and consultative support to formulate feasible adaptations, regular classroom teachers may be able to address the problems of some portion (in this case, 30%) of students who initially demonstrate significant learning discrepancies from classroom peers. Nevertheless, this database simultaneously indicates that some students will remain unresponsive to an adapted general education environment. This unresponsiveness creates the need for additional resources—specifically, the individualized instruction, the small-size instructional groups, and the more highly trained teachers available through special education—to address the learning problems of a small portion of learners.

Teacher Planning and Student Achievement in Special Education

Evidence supporting CBM's utility in helping special educators plan more effective programs is strong. Corroborating evidence (e.g., L. S. Fuchs, Deno, & Mirkin, 1984; L. S. Fuchs, et al., 1991a; L. S. Fuchs, et al., 1992; Jones & Krouse, 1988; Stecker & Fuchs, 1999; Wesson, Skiba, et al., 1984; Wesson, 1991) shows dramatic effects on student outcomes in reading, spelling, and math when special educators rely on CBM to inform instructional planning. To illustrate this database, we briefly describe one study in reading.

L. S. Fuchs et al. (1984) conducted a study in the New York City Public Schools. Teachers participated for 18 weeks in a contrast group or a CBM treatment group, where teachers conducted reading CBM at least twice weekly, scored and graphed those performances, and used prescriptive CBM decision rules for planning the students' reading programs. Children whose teachers employed CBM to develop reading programs achieved better than students whose teachers used conventional monitoring methods on the Passage Reading Test and on the decoding

and comprehension subtests of the Stanford Diagnostic Reading Test. Respective effect sizes were 1.18, .94, and .99. This suggests that, despite CBM's focus on text reading fluency, teachers planned better reading programs comprehensively to include foci on fluency, decoding, and comprehension.

Evaluating Treatment Effectiveness

To function adequately as a tool for evaluating treatment effectiveness, CBM must be able to answer questions such as the following: Is the regular classroom environment producing adequate growth? Do adaptations introduced in the regular classroom setting result in an improved growth rate? Does the provision of specialized services enhance student learning? To answer these treatment effectiveness questions, the assessment must demonstrate sensitivity to student growth and to relative treatment effects, and it must permit comparisons of the effectiveness of alternative service delivery options. CBM demonstrates these features.

Sensitivity to Academic Change

In an early study devoted to the issue of sensitivity to academic change, Marston, Fuchs, and Deno (1986) tested students on traditional, commercial achievement tests and on curriculum-based reading and written language measures early in October and 10 weeks later in December. CBM registered more student growth than did the traditional tests, suggesting greater sensitivity to student growth. In an operational replication, published as a second study in the same article, Marston et al. corroborated findings with pre- and posttestings 16 weeks apart on traditional, commercial reading achievement.

Other research has directly compared the sensitivity of CBM pre- and postperformance levels to that of CBM slopes. For example, while investigating the effects of a 3-week winter break on students' math performance, Allinder and Fuchs (1994) contrasted (a) comparison of CBM performance levels before and after the break with (b) comparison of pre- and postbreak slopes of progress. Results differed by type of analysis: Effects of the break were not demonstrated when performance level was assessed; examination of slopes, however, showed that students with positive prebreak trends were affected adversely by the school break, whereas students with negative prebreak trends were not.

Studies have also demonstrated that slopes based on ongoing CBM data reflect treatment effects more sensitively than traditional measures administered on a pre- and postbasis. L. S. Fuchs, et al. (1989b) showed that on the Stanford Achievement Test–reading comprehension subtest, administered to detect incremental change between two points in time, change scores of the treatment groups were not significantly different, and the effect size was a relatively low .36. By contrast, on CBM slope data, differences between groups achieved statistical significance and were associated with a larger effect size of .86 standard deviations. This pattern showing substantially larger effect sizes for CBM slope data has been corroborated in other treatment effectiveness research (e.g., L. S. Fuchs, et al., 1991; L. S. Fuchs, Fuchs, Hamlett, Phillips, & Bentz, 1994). Evidence therefore suggests that CBM slopes may be sensitive to student growth and to the relative effects of alternative treatments.

Comparing Student Progress under Alternative Service Delivery Options

Given a sensitive measurement system, however, questions remain about how that system functions when comparing student progress under alternative service delivery options. Two CBM studies illustrate this type of decision making. Marston (1987–1988) compared the relative effectiveness of regular and special education by analyzing slope on weekly CBM reading scores. An initial pool of 272 fourth, fifth, and sixth graders were selected for the yearlong study on the basis of performance at or below the 15th percentile on the Minneapolis Benchmark Test. The CBM reading performance of these 272 children was measured weekly. The 11 students who (a) spent at least 10 weeks in regular education, (b) were referred to and placed in special education, and (c) spent at least 10 weeks in special education were the focus of the analysis.

To determine the relative treatment effects of the two service delivery arrangements, a repeated measures analysis of variance was applied to the CBM slope data. Slopes were significantly greater in special education than in regular education, with the average slopes increasing from .60 to 1.15 words across the two service delivery settings. For 10 of 11 students, slopes were larger in special education; in 7 of the 10 cases, the difference was dramatic.

In a similar way, D. Fuchs, Fuchs, and Fernstrom (1993) used slope to examine the relative effectiveness of special and regular education for individual students as they moved in the opposite direction: as they reintegrated into general education classrooms. Twenty-one special education

students had been randomly assigned to a condition designed to facilitate successful reintegration to regular classroom math instruction through a deliberate and systematic process involving transenvironmental programming and CBM. Special educators used CBM to inform and strengthen their planning in the area of mathematics; at the same time, they monitored the target students' CBM growth and that of three low-performing (but legitimate academic) members of the general education setting. When the target students' performance level approached that of the low-performing peers, reintegration occurred, and the onus for instruction was transferred to the regular classroom teacher. After reintegration, CBM data continued to be collected for the target student and for the low-performing peers in the regular classroom.

Within special education, the experimental students' slopes were significantly greater than that of the low-performing peers. However, after reintegration, the slopes of the target students plunged and were significantly lower than that of the comparison students. On average, 63% of the reintegrated students' CBM data points in regular education fell below trend lines that had been projected on the basis of their growth rates within special education. This compared to only 44% for the comparison peers, and represented a statistically significant difference. As with the Marston (1987–1988) study, this database clearly revealed the effectiveness of the special, over the general, education setting for many (although not all) students. Both studies demonstrate CBM's capacity to document the effects of service delivery options.

Using CBM to Identify Nonresponders

The traditional assessment framework for identifying students with learning disabilities relies on discrepancies between intelligence and achievement tests to operationalize unexpected underachievement. This traditional framework has been scrutinized and attacked due to measurement and conceptual difficulties. An alternative framework for identifying students with learning disabilities that has received increasing interest (see, e.g., Fuchs & Fuchs, 1998), is one in which learning disability is conceptualized as nonresponsiveness to otherwise effective instruction. It requires that special education be considered only when a child's performance reveals a dual discrepancy: The student not only performs below the level demonstrated by classroom peers but also demonstrates a learning rate substantially below that of classmates.

To illustrate the rationale for this focus on dual discrepancies, we have relied on the following example from pediatric medicine. The endocrinol-

ogist monitoring a child's physical growth is interested not only in height but also growth velocity over time (Rosenfeld, 1982). Given a child whose current height places him or her below the third percentile, the endocrinologist considers the possibility of underlying pathology and the need to intervene only if, in response to an adequately nurturing environment, the individual's growth trajectory is flatter than that of appropriate comparison groups. Based on long-term, large-scale normative information (Tanner & Davies, 1985), this criterion typically is operationalized at age 7 as an annual growth rate of less than 4 cm. Consequently, the physician judges the 7-year-old who manifests a large discrepancy in height status, but who is nonetheless growing at least 4 cm annually in response to a nurturing environment, to be deriving available benefits from that environment and to be an inappropriate candidate for special intervention.

The endocrinologist's decision-making framework reflects three assumptions. The first assumption is that genetic variations underlie normal development, producing a range of heights across the population. The second assumption is that in response to a nurturing environment, a short but growing child does not present a pathological profile indicative of a need for special treatment to produce growth. Instead, this profile suggests an individual who may legitimately represent the lower end of the normal distribution on height—an individual whose development is commensurate with his or her capacity to grow. The third assumption is that under these circumstances, special intervention is unlikely to increase adult height sufficiently to warrant the risks associated with that intervention. Of course, when questions about the quality of the environment arise, the response is to remove those uncertainties by enhancing nurturance, even with hospitalization (Wolraich, 1996), so growth can be tested under adequate environmental conditions.

Applied to education, this decision-making framework translates into three related propositions. First, because student capacity varies, educational outcomes will differ across the population of learners, and a low-performing child, who is nonetheless learning, may ultimately perform not as well as his or her peers. For example, we do not expect all children to achieve the same degree of reading competence: Some will become scholars of great literature; others will achieve the minimal levels of competence to permit satisfactory employment and successful parenting.

Second, if a low-performing child is learning at a rate similar to the growth rates of other children in the same classroom environment, he or she is demonstrating the capacity to profit from that educational environment. Additional intervention, therefore, is unwarranted, even though a discrepancy in performance level may exist. That is, given the benefits being derived from the classroom instructional environment, the student

probably does not require a unique form of instruction and probably is achieving commensurate with his or her capacity to learn. Moreover, the risks and costs associated with entering the special education system are deemed inappropriate and unnecessary because it is unlikely, in light of the growth already occurring, that a different long-term educational outcome could be achieved as a function of that intervention. Of course, the converse is also assumed. When a low-performing child is not manifesting growth in a situation where others are thriving, consideration of special intervention is warranted. Alternative instructional methods must be tested to address the apparent mismatch between the student's learning requirements and those represented in the conventional instructional program.

The third assumption is that when the vast majority of students in a classroom are achieving inadequate growth rates (in comparison to local or national norms), one must question the adequacy of that educational environment before formulating decisions about individual student responsiveness. In that case, classroom intervention aimed at enhancing the overall quality of the instructional program must occur. Growth under more nurturing environmental conditions must be indexed before any child's need for special intervention can be assessed.

Dual discrepancy as an index of *failure to thrive* has considerable intuitive and empirical appeal. This index deals directly with the problem at hand (e.g., poor reading, poor math skills), reflects a dynamic rather than a static approach to learning and assessment, and is data based. CBM is a promising tool for indexing treatment responsiveness due to its capacity to model student growth, to evaluate treatment effects, and to simultaneously inform instructional programming; however, such an identification model is more labor intensive than a traditional framework. It requires (a) assessment of every child in every classroom every week, (b) evaluation of progress on a regular basis, (c) formulating interventions within general education classrooms for children identified as dually discrepant, (d) implementing the interventions with fidelity, and (e) evaluating the effects of the intervention. The question becomes a treatment validity model worth these additional requirements. One way of addressing this question is through empirical comparison with other identification procedures. Speece and her colleagues (Speece & Case, 2001; Speece, Molloy, & Case, 2000) compared the dual discrepancy method to IQ-reading achievement discrepancy and low reading achievement definitions of reading disability in an epidemiological sample of first- and second-grade children to assess the validity of the dual discrepancy approach. The determination of dual discrepancy status was based on CBM collected across 6 months of a school year. The population ($n = 694$) was screened on CBM letter sound fluency (first grade) and reading aloud (second grade) to identify at-risk

(n = 144) and comparison (n = 129) samples. The at-risk children comprised the lowest 25% of children in each class. To form the comparison sample, five children representing a range of skill above the 30th percentile on the screening measures were selected from each class. The at-risk and comparison samples were followed throughout the school year and were administered a minimum of 10 CBM text reading probes to determine dual discrepancy status (CBM–DD). Other measures of intelligence and reading achievement were used to form the IQ-reading achievement discrepancy (IQ–DS) and low reading achievement (LA) groups. Validation measures included phonological processing variables, teachers' ratings of academic competence, problem behaviors, and social skills, age, gender, and race.

The poor reader groups were formed hierarchically. All children who exhibited a dual discrepancy were assigned to the CBM–DD group. Then the remaining children who met the classification criteria were assigned to the IQ–DS group for the first set of comparisons and then to the LA group for the second, and parallel, set of comparisons. Using these procedures, 47 children were identified as CBM–DD, 17 as IQ–DS, and 28 as LA. The prevalence of CBM–DD was 8.1% compared to 5.9% for IQ–DS (based on all children who met the classification criteria regardless of whether they qualified for either of the other groups). By definition, the LA group would reflect approximately 25% of the population because a standard score below 90 on a norm-referenced measure was used as the criterion. Thus, it appeared that the CBM–DD procedures are likely to identify a reasonable number of children as learning disabled assuming that follow-up prevention activities are implemented to reduce false positive intervention.

Importantly, the CBM–DD group exhibited significantly lower reading scores compared to the at-risk and comparison sample children not classified as poor readers. Also, the CBM–DD group, compared to IQ–DS, was more impaired on every measure with the exception of reading and word reading efficiency. There were fewer and more modest differences between CBM–DD and LA. Notably, the CBM–DD group had poorer phonological awareness skills at second grade and were rated lower on academic competence, problem behaviors, and social skills. Interestingly, the four cross-sectional comparisons that used grade (including IQ–DS) suggested that CBM–DD children become more impaired over time. It was also interesting that the CBM–DD group was younger than either of the other poor reader groups, suggesting that dual discrepancy may support early identification and intervention. In addition, the CBM–DD and LA groups each had racial distributions that reflected the proportions of majority and minority children in the schools, whereas the IQ–DS group had a disproportionately high number of majority children.

Consequently, these group comparisons support the construct validity: The CBM–DD group demonstrated more problems in skills that underlie beginning reading, and teachers viewed CBM–DD children as less academically competent and as exhibiting more social and classroom problem behaviors. Further, the dual discrepancy method demonstrated social consequential validity in that decisions to identify children in this manner reflected racial equity. Also, identified children were younger, a consequence valued by educators, and considerably fewer children were identified compared to a low achievement definition. The CBM–DD classification was the only method that reflected favorably on racial equity as well as early identification. Therefore, although additional work clearly is warranted, a treatment validity framework for identifying learning disabilities, using CBM as a measurement tool, represents a promising alternative.

Conclusion

CBM development began in 1972, with a small number of doctoral students working under the direction of Stanley L. Deno at the University of Minnesota, to develop an efficient assessment teachers could use to describe academic competence and growth and to plan more effective instruction. As illustrated by the research we summarized in this article, over the past 2 decades, the number of independent researchers studying CBM has grown as have the applications investigated. At present, CBM appears useful for describing competence, enhancing outcomes, evaluating treatment effects, and identifying treatment nonresponders. Additional research, exploring other uses and providing additional replication, is required, but CBM's potential has been persuasively demonstrated.

References

Algozzine, B., Christenson, S., & Ysseldyke, J. E. (1982). Probabilities associated with the referral to placement process. *Teacher Education and Special Education, 5*(3), 19–23.

Allinder, R. M., & Fuchs, L. S. (1994). Alternative ways of analyzing effects of a short school break on students with and without disabilities. *School Psychology Quarterly, 9,* 145–160.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52,* 219–232.

Deno, S. L., & Fuchs, L. S. (1987). Developing curriculum-based measurement systems for data-based special education problem solving. *Focus on Exceptional Children, 19*(8), 1–16.

Deno, S. L., Mirkin, P., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children, 49,* 36–45.

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1994). The measurement of change: Assessing behavior over time and within a developmental

context. In Lyon, G. R. (Ed.), *Frames of reference for the assessment of learning disabilities: New views on measurement issues* (pp. 29–58). Baltimore: Brookes.

Fuchs, D., Fuchs, L. S., & Fernstrom, P. J. (1993). A conservative approach to special education reform: Mainstreaming through transenvironmental programming and curriculum-based measurement. *American Educational Research Journal, 30*, 149–178.

Fuchs, D., Fuchs, L. S., Mathes, P., & Simmons, D. (1997). Peer-assisted learning strategies: Making classrooms more responsible to student diversity. *American Educational Research Journal, 34*, 174–206.

Fuchs, D., Fuchs, L. S., Thompson, A., Al-Otaiba, S., Yen, L., Yang, N., Braun, M., & O'Connor, R. (2001). Is reading important in reading-readiness program? A randomized field trial with teachers as program implementers. *Journal of Educational Psychology, 93*, 251–267.

Fuchs, L. S. (1993). Enhancing instructional programming and student achievement with curriculum-based measurement. In J. J. Kramer (Ed.), *Curriculum-based assessment* (pp. 65–104). Lincoln: Buros Institute of Mental Measurements, University of Nebraska.

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children, 57*, 488–501.

Fuchs, L. S., & Deno, S. L. (1992). Effects of curriculum within curriculum-based measurement. *Exceptional Children, 58*, 232–243.

Fuchs, L. S., Deno, S. L., & Marston, D. (1983). Improving the reliability of curriculum-based measures of academic skills for psychoeducational decision making. *Diagnostique, 8*, 135–149.

Fuchs, L. S., Deno, S. L., & Mirkin, P. K. (1984). The effects of frequent curriculum-based measurement and evaluation on student achievement, pedagogy, and student awareness of learning. *American Educational Research Journal, 21*, 449–460.

Fuchs, L. S., & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psychology Review, 21*, 45–58.

Fuchs, L. S., & Fuchs, D. (1998). Curriculum-based measurement: A unifying framework for conceptualizing learning disability. *Learning Disability Research and Practice, 13*, 204–219.

Fuchs, L. S., & Fuchs, D. (1999). Monitoring student progress toward the development of reading competence: A review of three forms of classroom-based assessment. *School Psychology Review, 28*, 659–671.

Fuchs, L. S., Fuchs, D., & Deno, S. L. (1982). Reliability and validity of curriculum-based informal reading inventories. *Reading Research Quarterly, 18*, 6–26.

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1989a). Effects of alternative goal structures within curriculum-based measurement. *Exceptional Children, 55*, 429–438.

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1989b). Effects of instrumental use of curriculum-based measurement to enhance instructional programs. *Remedial and Special Education, 10*(2), 43–52.

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1989c). Monitoring reading growth using student recalls: Effects of two teacher feedback systems. *Journal of Educational Research, 83*, 103–111.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Allinder, R. M. (1989). The reliability and validity of skills analysis within curriculum-based measurement. *Diagnostique, 14*, 203–221.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Allinder, R. M. (1991a). Effects of expert system advice within curriculum-based measurement on teacher planning and student achievement in spelling. *School Psychology Review, 20*, 49–66.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Ferguson, C. (1992). Effects of expert system consultation within curriculum-based measurement using a reading maze task. *Exceptional Children, 58*, 436–450.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N. B., & Bentz, J. (1994). Classwide curriculum-based measurement: Helping general educators meet the challenge of student diversity. *Exceptional Children, 60*, 518–537.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N. B., & Karns, K. (1995). General educators' specialized adaptation for students with learning disabilities. *Exceptional Children, 61*, 440–459.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N., Karns, K., & Dutka, S. (1997). Enhancing students' helping behavior during peer-mediated instruction with conceptual mathematics explanations. *Elementary School Journal, 97*, 223–250.

Fuchs, L.S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1991). Effects of curriculum-based measurement and consultation on teacher planning and student achievement in mathematics operations. *American Educational Research Journal, 28*, 617–641.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Thompson, A., Roberts, P. H., Kubec, P., & Stecker, P. M. (1994). Technical features of a mathematics concepts and applications curriculum-based measurement system. *Diagnostique, 19*(4), 23–49.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review, 22*, 27–48.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education, 9*(2), 20–29.

Fuchs, L. S., & Hamlett, C. L. (1997). *Predictive validity of curriculum-based measurement with respect to the TerraNova*. Unpublished data.

Fuchs, L. S., Hosp, M., Fuchs, D., & Hamlett, C. L. (2001a). *Effects of diagnostic feedback within curriculum-based measurement*. Manuscript in preparation.

Fuchs, L. S., Hosp, M., Fuchs, D., & Hamlett, C. L. (2001b). *Psychometric features of the Peabody Phonics Inventory*. Unpublished manuscript.

Good, R. H., Deno, S. L., & Fuchs, L. S. (1995, February). *Modeling academic growth for students with and without disabilities*. Paper presented at the third annual Pacific Coast Research Conference, Laguna Beach, CA.

Good, R. H., & Shinn, M. R. (1990). Forecasting accuracy of slope estimates for reading curriculum-based measurement: Empirical evidence. *Behavioral Assessment, 12*, 179–194.

Hosp, M., & Fuchs, L. S. (2001). *Technical features of curriculum-based measurement's reading aloud task in the early grades*. Manuscript submitted for publication.

Jones, E. D., & Krouse, J. P. (1988). The effectiveness of data-based instruction by student teachers in classrooms for pupils with mild learning handicaps. *Teacher Education and Special Education, 11*, 9–19.

Marston, D. (1987–1988). The effectiveness of special education: A time-series analysis of reading performance in regular and special education settings. *The Journal of Special Education, 21*(4), 13–26.

Marston, D. (1989). Curriculum-based measurement: What is it and why do it? In M. R. Shinn (Eds.). *Curriculum-based measurement: Assessing special children* (pp. 18–78). New York: Guilford.

Marston, D., & Deno, S. L. (1981). *The reliability of simple, direct measures of written expression* (Research Report No. 50). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Marston, D., Fuchs, L. S., & Deno, S. L. (1986). Measuring pupil progress: A comparison of standardized achievement tests and curriculum-related measures. *Diagnostique, 11*, 71–90.

Marston, D., & Magnusson, D. (1988). Curriculum-based assessment: District-level implementation. In J. Graden, J. Zins, & M. Curtis (Eds.), *Alternative educational delivery systems: Enhancing instructional options for all students* (pp. 137–172). Washington, DC: National Association of School Psychologists.

Rosenfeld, R. G. (1982). Short stature. In M. Green and J. Haggerty (Eds.). *Ambulatory Pediatrics–IV*. Philadelphia: Saunders.

Shinn, M. R., Tindal, G., Spira, D., & Marston, D. (1987). Practice of learning disabilities as social policy. *Learning Disability Quarterly, 10*, 17–28.

Speece, D. L., & Case, L. P. (2001). Classification in context: An alternative approach to identifying early reading disability. *Journal of Educational Psychology, 93,* 735–749.

Speece, D. L., Molloy, D. E., & Case, L. P. (2000, February). *Toward validating a model of reading disability identification based on response to treatment.* Presented at the annual Pacific Coast Research Conference, La Jolla, CA.

Stecker, P. M., & Fuchs, L. S. (1999). Effecting superior achievement using curriculum-based measurement: The importance of individual progress monitoring. *Learning Disabilities Research and Practice, 15,* 128–134.

Tanner, J. M., & Davies, P. S. W. (1985). Clinical longitudinal standards for height and weight velocity for North American children. *Journal of Pediatrics, 107,* 317.

Wesson, C. L. (1991). Curriculum-based measurement and two models of follow-up consultation. *Exceptional Children, 57,* 246–257.

Wesson, C. L., Skiba, R., Sevcik, B., King, R., & Deno, S. (1984). The effects of technically adequate instructional data on achievement. *Remedial and Special Education, 5,* 17–22.

Willett, J. B., Ayoub, C. C., & Robinson, D. (1991). Using growth modeling to examine systematic differences in growth: An example of change in the functioning of families at risk of maladaptive parenting, child abuse, or neglect. *Journal of Consulting and Clinical Psychology, 59,* 38–47.

Wolraich, M. (Ed.). (1996). *Disorders of development and training: A practical guide to assessment and management* (2nd ed.). St. Louis: Mosby.